

## **APPROACHES TO DOCUMENTING PERFORMANCE CHARACTERISTICS OF BIOLOGICAL METHODS: THE PRECISION OF FIELD SAMPLING AND TAXONOMIC IDENTIFICATIONS**

### 1. Introduction

#### **Definition:**

##### *Data Quality Objectives (DQO)*

Statements of the level of uncertainty that a decision maker is willing to accept in the decisions made on the basis of the measurement data

#### **Definition:**

##### *Measurement Quality Objectives (MQO)*

Called data quality indicators by Keith (1988), MQOs are limits for the uncertainty of specific measurements (Keith 1991); they are statements that contain specific units of measure, such as percent recovery, percent relative standard deviation, standard deviation, root mean square error, etc.; also known as method performance characteristics; they should be thoroughly specified to allow direct comparisons among results.

Assessment and documentation of method performance characteristics is essential for appropriate application of environmental sampling and analysis methods, and interpretation of results. One of these characteristics, method precision, is important for establishing and evaluating measurement quality objectives (MQOs). Further, quantification of method precision is necessary to develop data quality objectives (DQOs) for program design and assessment precision. Similarly, the sensitivity of a method provides an indication of the responsiveness of an indicator to the stressor or stressors of concern. There are a variety of statistical methods that can be used to assess the precision and sensitivity of a method; this issue paper documents some of these approaches using case studies in which biological indicators were developed and tested. In particular, we investigate ways of assessing the precision associated with a particular method's ability to consistently measure individual properties of a biological assemblage. Several case studies are presented that demonstrate calculation of precision for field benthic macroinvertebrate sampling. We also show how precision of benthic macroinvertebrate taxonomy can be evaluated using independent taxonomists and laboratories. Data collection procedures and the statistical methods used to assess the precision and sensitivity of the particular method are described for each case study. Recommendations on precision assessment approaches are presented and the importance of precision calculations for quality assurance, DQOs, and management, are discussed.

### 2. Case Studies

#### **2.1 Precision of Field Sampling**

## 2.1.1 Maryland Department of Natural Resources

### **Benthic Macroinvertebrate Metrics used in Maryland DNR's Benthic IBI:**

Total Taxa (BIBI)

EPT Taxa

Ephemeroptera Taxa

Diptera Taxa

% EPT

% Ephemeroptera

% Tanytarsini

%Tanytarsini of Chironomidae

Becks Biotic Index

# Intolerant Taxa

% Tolerant

# Scraper Taxa

% Clingers

% Collector-Gatherers

This analysis is designed to test the precision of the methods being used in a statewide biological assessment program for Maryland streams and watersheds, the Maryland Biological Stream Survey (MBSS). Duplicate samples were collected at 27 100-meter stream reaches distributed throughout the study area. For this program, benthic macroinvertebrate sampling (in this case, the entire state) consists of using a D-frame net, and performing multihabitat collections over a 100-meter stream reach (Kazyak 1997). Sub-habitats (snags, riffles, undercut banks, root wads, macrophyte beds, leaf packs) are sampled in proportion to their frequency of occurrence within the reach. The leaf litter, substrate, or other detrital materials from twenty sweeps (or "jabs") are composited; a randomized, 100-organism subsample is sorted from each. Once metrics were calculated for the duplicate samples, two measures of precision were calculated: relative percent difference and root mean square error.

Relative percent differences (RPDs) were calculated for the metrics and index from the 27 sites using the equation:

where A is the metric or index value of the first sample and B is the metric or index value of the second sample. Relative percent difference represents precision as the difference between the duplicate metric values from each site. Lower RPDs indicate greater precision, therefore, metrics such as "EPT Taxa" and "Total Taxa", which had the lowest RPDs (medians of 17 and

18, respectively), were represented as having the greatest precision (Table 1). Metrics such as “%Tanytarsini of Chironomidae”, which had a median RPD of 67 percent, exhibited low precision in the two samples. Pairs of index scores from 20 out of the 27 sites had RPDs of less than 20 percent; nine of those had perfect agreement (RPD = 0) indicating that the index scores for the duplicate samples at those sites were identical. The median of the overall index RPDs was eight percent, indicating the index scores were more precise than any of the individual metrics which had medians that ranged from 16 to 67 percent (Table 1). Sampling precision, or consistency, was quantitatively characterized as RPD = 8, based on the overall benthic IBI score. It should be noted that the RPD statistic is inappropriate when one sample contains parameter

**Table 1.** Relative percent differences (RPDs) calculated from 27 duplicate benthic macroinvertebrate sample pairs, nontidal stream reaches. Variables calculated are metrics; index score refers to overall, aggregated Index of Biotic Integrity (IBI).

Site pair	RPDs														Mean	Median	Minimum	Maximum	RPD Index score
	Tot Taxa	EPT Taxa	Ephem Taxa	Dip taxa	% EPT	% Ephem	% Tany	% Tany Chiron	% Tany of Becks	# Intol taxa	% Tolerant taxa	# Scrap taxa	% Clingers	% Coll-Gath					
1	32	67	0	11	24	0	17	29	0	40	2	0	7	87	23	14	0	87	0
2	11	20	22	15	2	53	200	200	18	29	22	0	34	37	47	22	0	200	7
3	17	17	67	35	28	64	200	200	17	100	70	0	23	39	63	37	0	200	41
4	0	15	50	80	0	6	0	0	0	0	10	0	13	31	15	3	0	80	9
5	15	0	40	32	45	55	200	200	0	0	21	0	39	71	51	35	0	200	33
6	35	0	40	86	51	115	0	0	111	120	28	67	2	107	54	45	0	120	33
7	26	0	0	25	0	0	6	15	0	0	1	0	0	2	7	0	0	26	17
8	14	29	0	40	57	19	113	111	0	40	32	200	81	32	55	36	0	200	14
9	29	0	0	13	3	6	0	0	24	0	46	67	28	29	19	17	0	67	0
10	29	0	200	40	46	200	0	0	67	0	26	67	157	200	74	43	0	200	36
11	5	0	0	46	8	26	200	200	35	50	50	29	5	64	51	32	0	200	0
12	14	0	40	50	27	35	0	0	35	18	73	200	73	77	46	35	0	200	24
13	43	46	15	32	33	5	1	40	30	48	66	29	17	30	31	31	1	66	21
14	35	32	40	10	10	95	101	107	73	57	31	120	1	52	54	46	1	120	0
15	23	16	18	11	2	4	60	0	5	17	152	111	2	30	32	16	0	152	5
16	22	7	22	100	10	35	7	67	8	22	71	86	13	45	37	22	7	100	0
17	22	33	22	40	9	8	200	200	0	12	45	40	21	18	48	22	0	200	23
18	17	13	15	25	4	7	200	200	31	11	38	22	24	4	44	19	4	200	0
19	8	10	0	0	38	49	47	83	22	40	64	22	37	18	31	30	0	83	6
20	18	29	29	52	88	93	200	200	33	0	71	0	72	16	64	43	0	200	0
21	24	15	22	0	49	97	200	200	32	22	37	100	113	3	65	34	0	200	7
22	15	75	67	11	78	49	0	0	75	111	105	40	55	21	50	52	0	111	15
23	8	11	40	26	27	67	97	43	11	15	50	40	20	11	33	26	8	97	12
24	13	86	200	33	11	200	200	200	55	67	32	200	49	8	97	61	8	200	0
25	19	22	0	29	18	0	0	0	0	40	16	67	12	7	16	14	0	67	0
26	16	120	67	8	61	7	58	27	0	0	14	200	1	28	43	21	0	200	10
27	27	57	29	40	6	72	200	200	36	57	4	29	5	82	60	38	4	200	8
avg	20	27	39	33	27	51	93	93	27	34	44	64	34	43					12
median	18	17	22	32	24	35	60	67	22	22	37	40	21	30					8
min	0	0	0	0	0	0	0	0	0	0	1	0	0	3					0
max	43	120	200	100	88	200	200	200	111	120	152	200	157	200					41

values of 0 and the other contains values of 1 or greater because, in this case, RPD values can be as high as 200 percent. The metric “% Tanytarsini in Chironomidae” had an average RPD of 93 percent with numerous individual site values of 200 percent because of the absence of Tanytarsini at some sites. Values this high, which result from as little as a one unit difference between samples, should not be regarded as indicative of sampling consistency.

The root mean square error (RMSE) for the 27 duplicate sample pairs was calculated as an additional measure of precision (Table 2). The RMSE, which assumes normal distribution of the calculated metric values, is calculated by performing an ANOVA on the duplicates that produces a mean square error value (MSE) which represents the “within-group” variance. The square root of this variance is the estimated population standard deviation, or RMSE. It presents precision as the range within which the sample mean is likely to fall. The narrower the range the greater is the likelihood that the sample mean is representative of the true mean (i.e., the narrower the range, the greater the precision). Precision increases as the RMSE decreases, therefore, metrics such as “Ephemeroptera taxa” (RMSE = 0.9) exhibited greater precision than metrics such as “%Tanytarsini of Chironomidae” (RMSE = 11.7) (Table 2). The low RMSE value of the “Ephemeroptera taxa” metric indicates that this metric was similar in both samples, while the high RMSE of the “%Tanytarsini of Chironomidae” metric indicates that this metric differed between the two samples. Similar to the RPD precision estimates, RMSE calculations indicated the index score was more precise than the individual metrics (Figure 1). The metric RMSEs ranged between 0.9 and 11.7, while the overall index score RMSE was 0.4.

**Table 2.** Analysis of variance (ANOVA) with metric values and index score as dependent variables; sites as independent. Each site (n = 27) has one sample pair.

Metric <sup>a</sup>	RMSE	Approx. Mean <sup>b</sup>	Approx. Coefficient of Variation (%)	Detectable Differences (p = 0.10)		
				Single Sample	Duplicate Samples	Triplicate Samples
Total taxa	3.1	18.9	16.1	5.08	3.59	2.94
EPT taxa	2.1	8.6	24.7	3.44	2.44	1.99
Ephem taxa	0.9	3.3	27.4	1.48	1.04	0.85
Diptera taxa	1.8	8.0	22.7	2.95	2.09	1.70
% Ephem	6.8	19.7	34.8	11.15	7.89	6.44
% Tanytarsini	2.0	2.9	68.6	3.28	2.32	1.89
% Tany of Chir	11.7	13.6	86.2	19.19	13.57	11.08
Becks Biotic Index	2.2	8.7	24.6	3.61	2.55	2.08
Intolerant taxa	1.6	5.2	30.8	2.62	1.86	1.51
% Tolerants	11.4	37.9	30.2	18.70	13.22	10.79
Scraper taxa	1.3	2.4	53.3	2.13	1.51	1.23
% Clingers	10.5	47.2	22.2	17.22	12.18	9.94
% Collector	7.6	30.9	24.5	12.46	8.81	7.20

Index Score	0.4	3.02	11.9	0.66	0.46	0.38
-------------	-----	------	------	------	------	------

**2.1.2 Prince George’s County, Maryland, Department of Environmental Resources**

This study assessed the ecological condition of 55 stream sites in Prince George’s County using benthic macroinvertebrate, fish, physical habitat, and water chemistry data collected during the Spring 1999 index period (March 1-30). Field sampling and data analysis methods were used that are based on U.S. Environmental Protection Agency (U.S. EPA) Rapid Bioassessment Protocols (RBP) and enhance the comparability of results with adjacent county programs, the Maryland Biological Stream Survey, the State of Delaware, U.S. EPA regional assessments, and national efforts at establishing biological criteria for water resource protection. Benthic macroinvertebrates were collected from 100 meter (non-tidal stream) reaches by making one-meter linear sweeps (jabs) with a D-frame net (600 □ mesh) through different habitat types sampled in proportion to their frequency of occurrence in each reach. In the lab, the composited samples were randomly subsampled to 100 organisms, identified to genus, and metrics and an overall index were calculated. Duplicate samples were taken at six sites (approximately 10% of total sites sampled) to estimate precision using RPDs and RMSEs. However, unlike the previous study that tested precision using duplicate samples taken from the same reach, this study used duplicates from adjacent reaches to reduce the possibility of repeat sampling of the same habitat areas within a reach. The RPDs, which estimated the difference of IBI scores between sample pairs, were calculated for five of the six duplicate samples (one duplicate contained too few specimens) and ranged from 14 to 23 percent (Table 3). The RMSE for six metrics ranged from 0.7 to 8.8, but similar to the study discussed previously, the overall composite index score (IBI) was more precise (0.3) than any of the individual metrics (Table 4).

**Table 3.** Relative percent difference (RPD) for benthic IBI scores.

Station ID	Sample 1	Sample 2	RPD
09-009	IBI not calculated, too few organisms		
14-001	1.86	1.57	16.67
18-006	2.71	2.14	23.53
27-057	2.43	3.00	21.05
42-010	3.57	3.00	17.39
42-020	1.86	2.14	14.29

**Table 4.** Sampling precision estimates (root mean square error [RMSE]) derived from

ANOVA calculated on duplicate benthic macroinvertebrate samples (n = 5 sample pairs).

Metric/Index Score	Detectable Differences (p=0.10)			
	RMSE	Single Sample	Duplicate Samples	Triplicate Samples
Total Taxa	2.5	4.1	2.9	2.4
EPT Taxa	2.0	3.3	2.3	1.9
%Ephemeroptera	1.2	2.0	1.4	1.2
%Tanytarsini of Chironomidae	3.1	5.1	3.5	2.9
Beck's Biotic Index	2.8	4.7	3.3	2.7
No. Scraper Taxa	0.7	1.2	0.8	0.7
%Clingers	8.8	14.5	10.2	8.4
Index Score	0.3	0.4	0.3	0.2

### 2.1.3 Wyoming Department of Environmental Quality

#### **Benthic Macroinvertebrate Metrics used in Wyoming's Stream Integrity Index (WSII):**

- Total Taxa
- Ephemeroptera Taxa
- Plecoptera Taxa
- Trichoptera Taxa
- % Non-Chironomid Diptera
- % Non-insects
- % Hilsenhoff Biotic Index
- % 5 Dominants
- % Scrapers
- % Tolerants

This study was initiated to develop biological and physical habitat indicators for assessing the impairment status of streams in Wyoming. The Wyoming DEQ sampled benthic macroinvertebrates, physical habitat quality, selected field chemistry, and fluvial geomorphologic data from approximately 300 stream sites over a four-year period. To allow assessment of precision, duplicate benthic samples were taken at 25 streams. WDEQ's sampling method uses a square-foot surber sampler, takes eight (8) 1 ft<sup>2</sup> substrate samples from riffles, and composites them into a single sample. A randomized 500-organism subsample is taken from each, identified

to species level, and a series of metrics calculated (box at right). Sample duplicates were taken from different riffle areas of the same 100-meter stream reach. The RMSE as described above, was calculated for each of the nine metrics and the overall index, the Wyoming Stream Integrity Index (WSII) (Table 5). Unlike the previous case studies, the overall index here was not more precise than the metrics; five of the nine metrics had lower RMSEs (0.27-1.3) than the overall index (2.0). The “detectable difference” values represent the range that a particular metric is likely to fall within 90 percent of the time (i.e., the 90% confidence interval). Metrics that have higher RMSEs have wider confidence intervals because they are less precise and less likely to fall within a narrow range of values for each replicate sampling.

**Table 5.** Statistics of repeated samples in Wyoming and the detectable difference (effect size) at 0.10 significance level. The index is on a 100 point scale.

Metric <sup>a</sup>	RMSE	Approx. Mean <sup>b</sup>	Approx. Coefficient of Variation (%)	Detectable Differences (p = 0.10)		
				Single Sample	Duplicate Samples	Triplicate Samples
Total Taxa	4.1	35.9	11.5	7 taxa	5 taxa	5 taxa
Ephemeroptera taxa	0.9	6.8	13.3	2 taxa	1 taxon	1 taxon
Plecoptera taxa	1.0	4.8	21.2	2 taxa	1 taxon	1 taxon
Trichoptera taxa	1.1	6.9	15.3	2 taxa	1 taxon	1 taxon
% non-insects	3.8	8.9	42.9	6.3 %	4.4 %	4.3 %
% Diptera (non-chironomid)	1.3	5.1	25.0	2.1 %	1.5 %	1.4 %
HBI	0.27	3.43	7.85	0.44 units	0.31 units	0.26 units
% 5 dominant taxa	4.3	64.2	6.7	7.1 %	5.0 %	4.1 %
% scrapers	4.8	25.5	18.9	7.9 %	5.6 %	4.6 %
Overall index	2.0	70.0	2.9	3.3 units	2.3 units	1.9 units

<sup>a</sup> Percent tolerant metric not included in duplicate site calculations.

<sup>b</sup> Mean of 25 replicated sites; population means may differ.

### 2.1.4 Arizona Department of Environmental Quality

The objective of this study was to develop a biological index for warmwater streams in Arizona; to determine the relative advantages of identifying organisms to genus or family; to determine whether both fall and spring sample collection are necessary; and to determine whether single habitat or multiple habitat sampling is best. Benthic macroinvertebrate data were collected by the Arizona Department of Environmental Quality (AZDEQ) from 1992-1995 during both the spring and fall seasons and in both riffle and pool habitats using modified RBPs. The Arizona sampling protocol (Meyerhoff and Spindler 1994) consists of composited 1-minute kick samples collected from 3 riffle and 3 pool habitats within the sample reach, using a D-frame net.

Microhabitats were sampled separately by hand-picking organisms from microenvironments for 30 minutes. A minimum of 300 macroinvertebrates were randomly subsampled and identified to the lowest practical taxonomic level, usually genus (Meyerhoff and Spindler 1994) and metrics

and an overall index were calculated. Unlike the other studies discussed here that used the RMSE of duplicate samples to estimate precision, this study used RMSEs of measurements taken over several years, during different seasons, and in different habitats to estimate the precision of the method. All multiple observations of each metric at single sites were used as replicate observations in the ANOVA, with the site as the primary treatment variable. The different replicates yielded similar RMSEs that ranged from 6 to 9 indicating that precision estimates were similar for multiple observations among seasons, habitats, and years (Table 6). Before this study, Arizona had been conducting monitoring in two index periods and in two habitats, but because of the replicability of the data as reported in this study, it was recommended that future sampling be conducted in only one index period and in only one habitat.

**Table 6.** Root mean-square errors of preliminary (aggregated) index score, from repeated observations of the same sites

<b>Sample Group Analyzed</b>	<b>Among</b>	<b>N</b>	<b>RMSE</b>
Spring Pool	years	21	7.64
Spring Riffle	years	22	6.52
95 Pool	seasons	34	7.44
95 Riffle	seasons	60	8.56
95 Fall	habitats	42	6.44
95 Spring	habitats	44	5.93

## 2.2 Precision of Taxonomic Identifications

This case study presents side-by-side comparisons of the taxonomic results from two different laboratories on the same benthic macroinvertebrate samples. In 1996, the Maryland DNR’s benthic taxonomy laboratory began re-identifying approximately 1,100 samples from family level to genus. Prior to that, all biological assessments had been performed using the higher level taxonomy (i.e., family). Following re-identification, the Maryland Biological Stream Survey (MBSS) determined a need to perform additional QA/QC on this portion of their sampling and analysis program, and that an assessment of taxonomic precision would be appropriate. From among the 1,100 samples, 55 (or 5%) were randomly selected. Vials and slides constituting each were reassembled, sent to a third party “assessor”, and then delivered to a second taxonomic laboratory. All samples were re-identified by the second laboratory, and the results entered on a spreadsheet.

For each of the samples, results were put side-by-side, and included taxonomic name and the

number of individuals of each (count). Table 7 shows, as an example, one of the sample comparisons. Direct comparison of individual sample results by independent taxonomists requires line by line examination by the third party.

Results showed that the percent taxonomic disagreement for the 55 samples ranged from 0 to 50 percent with an average of 17 percent disagreement (Figure 2, Table 8). As an example of QC “problems” encountered, there were several recurring disagreements including, but not limited to, the following genera:

- *Synurella* vs. *Crangonyx* (amphipods)
- *Oemopteryx* vs. *Strophopteryx* (stoneflies)
- *Allocapnia* vs. *Paracapnia* (stoneflies)
- *Fossaria* vs. *Pseudosuccinea* (snails)
- *Goniobasis* vs. *Pleurocera* (snails)

There was also relatively consistent disagreement on blackflies (Simuliidae), and their genus-level determination as *Simulium*, *Prosimulium*, or *Stegopterna*. Specimens of the stonefly (Plecoptera) families Nemouridae and Taeniopterygidae typically produced problematic identifications when they were of early instars, many of their morphological characters having not yet fully developed. These comparisons did not include insects of the family Chironomidae.

Following initial comparisons, discussions were held to determine if some consensus could be reached between the two laboratories. There was also concern about “double counting” disagreements; for example, if an individual organism was identified and counted as *Synurella* (Crustacea: Amphipoda) by one taxonomist and as *Crangonyx* (Crustacea: Amphipoda) by the other. This was counted as one disagreement, not two. In many cases, Laboratory 1 felt confident in placing a genus level name on an organism, while Laboratory 2 chose leaving a more coarse-level name (i.e., family) on it. There could be several reasons for this: damaged specimen with diagnostic morphological features missing (such as gills, legs, caudal filaments, or antennae) early instar (juvenile) specimens with diagnostic morphological features underdeveloped or antennae) early instar (juvenile) specimens with diagnostic morphological features underdeveloped or only minimally visible, and, experience of the individual taxonomist. This was not counted as a disagreement between taxonomists.

**Table 7.** Side-by-side comparison of the taxonomic results from two laboratories on the same sample.

	Site	Order +	Family	Lab No. 1 ID	Lab No. 1 Count	Lab No. 2 Count	Lab No. 2 ID	Comments/Results	Genus ID*
1	95078	Acerpenna (no head)	NO COUNT	Acerpenna (no head)	1	0		?	
2	95078	Coleoptera	Ptilodactylidae	Anchytarsus	1	1	Anchytarsus	ok	1
3	95078	Diptera	Ceratopogonidae		0	2	Bezzia	ok/mount vs no mount	1
4	95078	Diptera	Ceratopogonidae	Ceratopogonidae	2	0		ditto/see row no. 3	
5	95078	Diptera	Chironomidae		0	2	Conchapelopia	<b>difference in overall no. midge individuals (Lab 1: 27, Lab 2: 21)</b>	
6	95078	Diptera	Chironomidae		0	3	Paratrichocladius		
7	95078	Diptera	Chironomidae		0	1	Potthastia		
8	95078	Diptera	Chironomidae	Brillia	1	1	Brillia		
9	95078	Diptera	Chironomidae	Chironomidae	18	0			
10	95078	Diptera	Chironomidae	Conchapelopia genus group	1	0			
11	95078	Diptera	Chironomidae	Diamesa	1	0			
12	95078	Diptera	Chironomidae	Micropsectra	2	3	Micropsectra		
13	95078	Diptera	Chironomidae	Orthocladius	2	7	Orthocladius		
14	95078	Diptera	Chironomidae	Parametriocnemus	1	3	Parametriocnemus		
15	95078	Diptera	Chironomidae	Sympotthastia	1	1	Sympotthastia		
16	95078	Diptera	Tabanidae	Chrysops	3	3	Chrysops	ok	1
17	95078	Diptera	Tipulidae		0	3	Ormosia	see row No. 20	1
18	95078	Diptera	Tipulidae	Pseudolimnophila	1	1	Pseudolimnophila	ok	1
19	95078	Diptera	Tipulidae	Tipula	1	1	Tipula	ok	1
20	95078	Diptera	Tipulidae	unkn. tipulid	2	0		Lab 2 id'd as Ormosia, Lab 1 found 1 less specimen	
21	95078	Ephemeroptera	Ameletidae	Ameletus	1	1	Ameletus	ok	1
22	95078	Ephemeroptera	Baetidae	Acerpenna	12	12	Acerpenna	ok	1
23	95078	Ephemeroptera	Ephemerellidae	Ephemerella	10	11	Ephemerella	ok, one spec. disagreement at family level, see row no. 25	1
24	95078	Ephemeroptera	Ephemerellidae	Eurylophella	9	9	Eurylophella	ok	1
25	95078	Ephemeroptera	Leptophlebiidae	leptophlebiid (inc., EI)	1	0		disagreement at family level, see row no. 23	
26	95078	Odonata	Aeschnidae	Boyeria	1	1	Boyeria	ok	1
27	95078	Odonata	Calopterygidae	Calopteryx	1	1	Calopteryx	ok	1
28	95078	Odonata	Cordulegasteridae	Cordulegaster	1	1	Cordulegaster	ok	1
29	95078	Odonata	Corduliidae		0	1	Somatochlora	disagreement Helocordulia vs. Somatochlora	0
30	95078	Odonata	Corduliidae	prob. Helocordulia	1	0		ditto/see row no. 29	

	Site	Order +	Family	Lab No. 1 ID	Lab No. 1 Count	Lab No. 2 Count	Lab No. 2 ID	Comments/Results	Genus ID*
				(EI)					
31	95078	Odonata	Gomphidae		0	1	Stylogomphus	ok	1
32	95078	Odonata	Gomphidae	unid. gomphid (EI)	1	0		ditto/see row no. 31	
33	95078	Odonata	Macromiidae	Macromia	1	1	Macromia	ok	1
34	95078	Plecoptera	Chloroperlidae		0	54	Sweltsa	ok	1
35	95078	Plecoptera	Chloroperlidae	prob. Sweltsa (EI)	54	0		ditto/row no. 34	
36	95078	Plecoptera	Perlidae	Eccoptura	2	2	Eccoptura	ok	1
37	95078	Plecoptera	Perlidae	Isoperla	5	5	Isoperla	ok	1
38	95078	prob. Erioptera (no head)	NO COUNT	prob. Erioptera (no head)	1	0		?	
39	95078	Trichoptera	Hydropsychidae	Hydropsyche	1	1	Hydropsyche	ok	1
40	95078	Trichoptera	Limnephilidae	Pycnopsyche	1	1	Pycnopsyche	ok	1
41	95078	Trichoptera	Phrygaenidae	Ptilostomis	1	1	Ptilostomis	ok	1
42	95078	Trichoptera	Uenoidae	Neophylax	15	17	Neophylax	ok, Lab 1 found two less specimens	1

ID identification

EI early instar

\* 1 = agreement; 0 = disagreement; cell shaded = not compared; cell empty = comparison made in another row

### Predominant Reasons for Taxonomic Disagreements:

- Different technical literature
- Early instar/morphological underdevelopment
- Damaged specimens/missing diagnostic features
- “missing” specimens

There were also some differences in counts, that is, in the number of individuals per taxon as recorded by each laboratory. Although this component of benthic macroinvertebrate sample processing does not always cause a problem, it occasionally does. Consequently, it is worth noting some of the causes for miscounts, how miscounts can cause instances of taxonomic disagreement, and how they might be avoided.

The scenario is this: Laboratory 1 finds and identified a specimen; Laboratory 2 does not find the specimen. There are two reasons this could happen. First, Laboratory 2 may have found the specimen and placed it under an incorrect name, appearing on its final datasheet as if that taxon were missing (if it was the only individual representing that taxon found by Laboratory 1). This should be interpreted as a true misidentification or disagreement. Second, very small specimens,

**Table 8.** Rate of agreement between two taxonomic laboratories in identification of identical

samples. Samples are from the Maryland Biological Stream Survey (MBSS).

Sam ple	Genus Level		
	No. Agreem	Total No. Identifica	Perce nt
9500 7	12	15	80
9501 9	9	11	82
9503 4	9	12	75
9504 9	7	10	70
9507 8	22	23	96
9528 3	15	18	83
9531 2	20	21	95
9532 4	10	10	100
9540 5	10	11	91
9603 8	15	16	94
9606 3	9	12	75
9606 8	6	8	75
9608 7	7	12	58
9609 1	17	19	89
9609 8	8	8	100
9610 3	7	8	88
9610 6	12	18	67
9611 4	15	16	94

9612 3	3	6	50
9613 0	3	3	100
9615 0	12	19	63
9615 6	6	12	50
9618 6	6	8	75
9619 9	10	11	91
9620 9	14	16	88
9622 3	5	5	100
9626 6	16	17	94
9627 8	7	7	100
<b>Genus Level</b>			
<b>Sam ple ID</b>	<b>No. Agree ments</b>	<b>Total No. Identifica tions</b>	<b>Perce nt Agree ment</b>
9628 9	17		85
9633 7	8	20	62
9635 8	23	25	92
9637 9	7	9	78
9700 9	9	14	64
9701 7	15	19	79
9702 7	10	11	91
9703 7	12	13	92
9705 0	18	19	95
9706 6	19	20	95

9707 1	8	9	89
9709 1	8	9	89
9712 6	12	20	60
9715 5	2	2	100
9715 7	11	14	79
9717 9	6	7	86
9719 1	10	12	83
9720 3	7	10	70
9722 3	9	12	75
9722 6	10	12	83
9724 1	5	6	83
9724 2	4	6	67
9725 5	20	21	95
9726 1	3	3	100
9727 4	18	20	90
9727 8	16	17	94
9729 3	9	11	82
<b>TOT AL</b>	588	706	83

such as first or second instar midges, mayflies, caddisflies (or others), could be hidden among body parts of larger specimens. For example, small mites or tiny magflies could be lodged under wingpads, heavy setal areas, or within mouthparts when Laboratory 1 is processing, only to become separated and evident to Laboratory 2. Also, each time a sample is handled, the potential for specimen loss is increased. Solutions for each include increasing the caution of sample sorters and taxonomists in locating potentially hidden, small specimens; and minimizing the number of times a sample is handled.

Recommendations were made to the MBSS to perform this assessment every 2-3 years. It was also recommended that a MQO for taxonomic precision in the range of 10-12% would be reasonable, recognizing that 5% should be considered essentially perfect agreement for this analytic process.

### 3. Discussion

The examples presented in this paper demonstrate how calculations can be made to document precision for benthic macroinvertebrate field sampling and taxonomic identifications. It must be made clear, however, that the sampling precision values are, in themselves, subject to variability based on other components of the overall sampling and assessment protocol. They were calculated using biological metric values and overall index scores. Metrics and index scores are also directly affected by, for example, number of samples, degree of subsampling, taxonomic level of final identifications used in metric calculation, sources of tolerance values and functional feeding group designations, the individual metrics that are selected to be calculated, and numeric thresholds used for scoring or evaluating the metrics. As is seen from the evaluation of taxonomic precision, a certain amount of identification "error" in a biological monitoring program may be inevitable. But, the extent to which it effects the precision of the overall assessment results may also be inconsequential. To completely understand which, if any, of the inherent error sources reduce the usefulness of an assessment, the extent of variability that is introduced by each must be understood.

Although RPDs and RMSEs are both precision estimates they present the information in different forms and may be uniquely appropriate at different times. When one sample contains parameter values of 0 and the other contains values of 1 or greater, RPD values can be as high as 200 percent, which is not representative of sampling precision. In the Maryland DNR example, the metric "% Tanytarsini in Chironomidae" had an average RPD of 93 percent with numerous individual site values of 200 percent because of the absence of Tanytarsini. The RMSE value for the same data indicated much greater precision in this metric (11.7%). In this circumstance, the RMSE value was more representative of the precision because it is not affected as dramatically by low parameter values.

RPDs can be calculated for individual replicate samples, while calculation of RMSEs requires an ANOVA to be performed on data from multiple, duplicated samples. RPD is, thus, a very straightforward calculation of the nearness of two values. As such, it can provide a "red flag" for duplicate samples that are beyond a control limit of, e.g., 10%. Such a flag might allow a "Quality Assessor" to identify previously unrecognized sources of error resulting from field sampling or laboratory analytical processes. RMSE does not provide detailed information necessary to identifying data anomalies resulting from unique site, habitat, field crew, or other site-specific conditions. It does, however, give a robust characterization of the (site) population-level variability of a measure. It also, very easily, allows calculation of confidence intervals around metric values and index scores.

#### **4.0 Literature Cited**

Kazyak, P.F. 1997. Maryland Biological Stream Survey Sampling Manual. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division, Annapolis, MD. 71 pp. & Appendices A and B.

Keith, L.H. (Editor). 1988. Principles of Environmental Sampling. American Chemical Society, Washington, DC. 458 pp.

Keith, L.H. 1991. Environmental Sampling and Analysis. A Practical Guide. Lewis Publishers, Chelsea, MI. 143 pp.

Meyerhoff, R.D. and P.H. Spindler. 1994. Biological Sampling Protocols: Reference Site Selection and Sampling Methods. Arizona Department of Environmental Quality, Phoenix, AZ.